Εντολές

hdfs dfs -ls ../hbase/dataset/

hbase shell

hdfs dfs -cat ../hbase/dataset/yelp\_checkin.csv

scan 'USER12.YELPBUSINESS',{ COLUMNS => 'ATTRIBUTES: ACCEPTSINSURANCE ',LIMIT => 10}

scan 'USER12.YELPCHECKIN'

Φόρτωση των δεδομένων:

hbase org.apache.hadoop.hbase.mapreduce.ImportTsv '-Dimporttsv.separator=,' -Dimporttsv.columns=HBASE\_ROW\_KEY,BASE:NAME,BASE:NEIGHBORHOOD,BASE:ADDRESS,BASE:CITY,BASE:STATE,BASE:POSTALCODE,BASE:LATITUDE,BASE:LONGTITUDE,BASE:STARS,BASE:REVIEWCOUNT,BASE:ISOPEN,BASE:CATEGORIES USER12.YELPBUSINESS ../hbase/dataset/yelp\_business.csv

Φόρτωση των δεδομένων:

hbase org.apache.hadoop.hbase.mapreduce.ImportTsv '-Dimporttsv.separator=,' -Dimporttsv.columns=HBASE\_ROW\_KEY,ATTRIBUTES:ACCEPTSINSURANCE,ATTRIBUTES:BYAPPOINTMENTONLY,ATTRIBUTES:BUSINESSACCEPTSCREDITCARDS,ATTRIBUTES:BUSINESSPARKINGGARAGE,ATTRIBUTES:BUSINESSPARKINGSTREET,ATTRIBUTES:BUSINESSPARKINGVALIDATED,ATTRIBUTES:BUSINESSPARKINGLOT,ATTRIBUTES:BUSINESSPARKINGVALET,ATTRIBUTES:HAIRSPECIALIZESINCOLORING,ATTRIBUTES:HAIRSPECIALIZESINAFRICANAMERICAN,ATTRIBUTES:HAIRSPECIALIZESINCURLY,ATTRIBUTES:HAIRSPECIALIZESINPERMS,ATTRIBUTES:HAIRSPECIALIZESINKIDS,ATTRIBUTES:HAIRSPECIALIZESINEXTENSIONS,ATTRIBUTES:HAIRSPECIALIZESINASIAN,ATTRIBUTES:HAIRSPECIALIZESINSTRAIGHTPERMS,ATTRIBUTES:RESTAURANTSPRICERANGETWO,ATTRIBUTES:GOODFORKIDS,ATTRIBUTES:WHEELCHAIRACCESSIBLE,ATTRIBUTES:BIKEPARKING,ATTRIBUTES:ALCOHOL,ATTRIBUTES:HASTV,ATTRIBUTES:NOISELEVEL,ATTRIBUTES:RESTAURANTSATTIRE,ATTRIBUTES:MUSICDJ,ATTRIBUTES:MUSICBACKGROUNDMUSIC,ATTRIBUTES:MUSICNOMUSIC,ATTRIBUTES:MUSICKARAOKE,ATTRIBUTES:MUSICLIVE,ATTRIBUTES:MUSICVIDEO,ATTRIBUTES:MUSICJUKEBOX,ATTRIBUTES:AMBIENCEROMANTIC,ATTRIBUTES:AMBIENCEINTIMATE,ATTRIBUTES:AMBIENCECLASSY,ATTRIBUTES:AMBIENCEHIPSTER,ATTRIBUTES:AMBIENCEDIVEY,ATTRIBUTES:AMBIENCETOURISTY,ATTRIBUTES:AMBIENCETRENDY,ATTRIBUTES:AMBIENCEUPSCALE,ATTRIBUTES:AMBIENCECASUAL,ATTRIBUTES:RESTAURANTSGOODFORGROUPS,ATTRIBUTES:CATERS,ATTRIBUTES:WIFI,ATTRIBUTES:RESTAURANTSRESERVATIONS,ATTRIBUTES:RESTAURANTSTAKEOUT,ATTRIBUTES:HAPPYHOUR,ATTRIBUTES:GOODFORDANCING,ATTRIBUTES:RESTAURANTSTABLESERVICE,ATTRIBUTES:OUTDOORSEATING,ATTRIBUTES:RESTAURANTSDELIVERY,ATTRIBUTES:BESTNIGHTSMONDAY,ATTRIBUTES:BESTNIGHTSTUESDAY,ATTRIBUTES:BESTNIGHTSFRIDAY,ATTRIBUTES:BESTNIGHTSWEDNESDAY,ATTRIBUTES:BESTNIGHTSTHURSDAY,ATTRIBUTES:BESTNIGHTSSUNDAY,ATTRIBUTES:BESTNIGHTSSATURDAY,ATTRIBUTES:GOODFORMEALDESSERT,ATTRIBUTES:GOODFORMEALLATENIGHT,ATTRIBUTES:GOODFORMEALLUNCH,ATTRIBUTES:GOODFORMEALDINNER,ATTRIBUTES:GOODFORMEALBREAKFAST,ATTRIBUTES:GOODFORMEALBRUNCH,ATTRIBUTES:COATCHECK,ATTRIBUTES:SMOKING,ATTRIBUTES:DRIVETHRU,ATTRIBUTES:DOGSALLOWED,ATTRIBUTES:BUSINESSACCEPTSBITCOIN,ATTRIBUTES:OPEN24HOURS,ATTRIBUTES:BYOBCORKAGE,ATTRIBUTES:BYOB,ATTRIBUTES:CORKAGE,ATTRIBUTES:DIETARYRESTRICTIONSDAIRYFREE,ATTRIBUTES:DIETARYRESTRICTIONSGLUTENFREE,ATTRIBUTES:DIETARYRESTRICTIONSVEGAN,ATTRIBUTES:DIETARYRESTRICTIONSKOSHER,ATTRIBUTES:DIETARYRESTRICTIONSHALAL,ATTRIBUTES:DIETARYRESTRICTIONSSOYFREE,ATTRIBUTES:DIETARYRESTRICTIONSVEGETARIAN,ATTRIBUTES:AGESALLOWED,ATTRIBUTES:RESTAURANTSCOUNTERSERVICE USER12.YELPBUSINESS ../hbase/dataset/yelp\_business\_attributes.csv

Φόρτωση των δεδομένων:

hbase org.apache.hadoop.hbase.mapreduce.ImportTsv '-Dimporttsv.separator=,' -Dimporttsv.columns=HBASE\_ROW\_KEY,HOURS:MONDAY,HOURS:TUESDAY,HOURS:WEDNESDAY,HOURS:THURSDAY,HOURS:FRIDAY,HOURS:SATURDAY,HOURS:SUNDAY USER12.YELPBUSINESS ../hbase/dataset/yelp\_business\_hours.csv

Φόρτωση των δεδομένων:

hbase org.apache.hadoop.hbase.mapreduce.ImportTsv '-Dimporttsv.separator=,' -Dimporttsv.columns=HBASE\_ROW\_KEY,PERHOUR:BUSINESSID,PERHOUR:WEEKDAY,PERHOUR:HOUR,PERHOUR:CHECKIN USER12.YELPCHECKIN ../hbase/dataset/yelp\_checkin.csv

hadoop fs -ls ../user/hbase/dataset

create 'USER12.YELPBUSINESS', 'BASE', 'ATTRIBUTES', 'HOURS'

create 'USER12.YELPCHECKIN', 'PERHOUR'

SQL:

/usr/hdp/2.6.3.0-235/phoenix/bin/sqlline.py

View για το table USER12.YELPBUSINESS:

CREATE VIEW USER12. YELPBUSINESS ( ROWKEY VARCHAR PRIMARY KEY,BASE.NAME VARCHAR,BASE.NEIGHBORHOOD VARCHAR,BASE.ADDRESS VARCHAR,BASE.CITY VARCHAR,BASE.STATE VARCHAR,BASE.POSTALCODE VARCHAR,BASE.LATITUDE VARCHAR,BASE.LONGTITUDE VARCHAR,BASE.STARS VARCHAR,BASE.REVIEWCOUNT VARCHAR,BASE.ISOPEN VARCHAR,BASE.CATEGORIES VARCHAR, ATTRIBUTES.ACCEPTSINSURANCE VARCHAR,ATTRIBUTES.BYAPPOINTMENTONLY VARCHAR,ATTRIBUTES.BUSINESSACCEPTSCREDITCARDS VARCHAR,ATTRIBUTES.BUSINESSPARKINGGARAGE VARCHAR,ATTRIBUTES.BUSINESSPARKINGSTREET VARCHAR,ATTRIBUTES.BUSINESSPARKINGVALIDATED VARCHAR,ATTRIBUTES.BUSINESSPARKINGLOT VARCHAR,ATTRIBUTES.BUSINESSPARKINGVALET VARCHAR,ATTRIBUTES.HAIRSPECIALIZESINCOLORING VARCHAR,ATTRIBUTES.HAIRSPECIALIZESINAFRICANAMERICAN VARCHAR,ATTRIBUTES.HAIRSPECIALIZESINCURLY VARCHAR,ATTRIBUTES.HAIRSPECIALIZESINPERMS VARCHAR,ATTRIBUTES.HAIRSPECIALIZESINKIDS VARCHAR,ATTRIBUTES.HAIRSPECIALIZESINEXTENSIONS VARCHAR,ATTRIBUTES.HAIRSPECIALIZESINASIAN VARCHAR,ATTRIBUTES.HAIRSPECIALIZESINSTRAIGHTPERMS VARCHAR,ATTRIBUTES.RESTAURANTSPRICERANGETWO VARCHAR,ATTRIBUTES.GOODFORKIDS VARCHAR,ATTRIBUTES.WHEELCHAIRACCESSIBLE VARCHAR,ATTRIBUTES.BIKEPARKING VARCHAR,ATTRIBUTES.ALCOHOL VARCHAR,ATTRIBUTES.HASTV VARCHAR,ATTRIBUTES.NOISELEVEL VARCHAR,ATTRIBUTES.RESTAURANTSATTIRE VARCHAR,ATTRIBUTES.MUSICDJ VARCHAR,ATTRIBUTES.MUSICBACKGROUNDMUSIC VARCHAR,ATTRIBUTES.MUSICNOMUSIC VARCHAR,ATTRIBUTES.MUSICKARAOKE VARCHAR,ATTRIBUTES.MUSICLIVE VARCHAR,ATTRIBUTES.MUSICVIDEO VARCHAR,ATTRIBUTES.MUSICJUKEBOX VARCHAR,ATTRIBUTES.AMBIENCEROMANTIC VARCHAR,ATTRIBUTES.AMBIENCEINTIMATE VARCHAR,ATTRIBUTES.AMBIENCECLASSY VARCHAR,ATTRIBUTES.AMBIENCEHIPSTER VARCHAR,ATTRIBUTES.AMBIENCEDIVEY VARCHAR,ATTRIBUTES.AMBIENCETOURISTY VARCHAR,ATTRIBUTES.AMBIENCETRENDY VARCHAR,ATTRIBUTES.AMBIENCEUPSCALE VARCHAR,ATTRIBUTES.AMBIENCECASUAL VARCHAR,ATTRIBUTES.RESTAURANTSGOODFORGROUPS VARCHAR,ATTRIBUTES.CATERS VARCHAR,ATTRIBUTES.WIFI VARCHAR,ATTRIBUTES.RESTAURANTSRESERVATIONS VARCHAR,ATTRIBUTES.RESTAURANTSTAKEOUT VARCHAR,ATTRIBUTES.HAPPYHOUR VARCHAR,ATTRIBUTES.GOODFORDANCING VARCHAR,ATTRIBUTES.RESTAURANTSTABLESERVICE VARCHAR,ATTRIBUTES.OUTDOORSEATING VARCHAR,ATTRIBUTES.RESTAURANTSDELIVERY VARCHAR,ATTRIBUTES.BESTNIGHTSMONDAY VARCHAR,ATTRIBUTES.BESTNIGHTSTUESDAY VARCHAR,ATTRIBUTES.BESTNIGHTSFRIDAY VARCHAR,ATTRIBUTES.BESTNIGHTSWEDNESDAY VARCHAR,ATTRIBUTES.BESTNIGHTSTHURSDAY VARCHAR,ATTRIBUTES.BESTNIGHTSSUNDAY VARCHAR,ATTRIBUTES.BESTNIGHTSSATURDAY VARCHAR,ATTRIBUTES.GOODFORMEALDESSERT VARCHAR,ATTRIBUTES.GOODFORMEALLATENIGHT VARCHAR,ATTRIBUTES.GOODFORMEALLUNCH VARCHAR,ATTRIBUTES.GOODFORMEALDINNER VARCHAR,ATTRIBUTES.GOODFORMEALBREAKFAST VARCHAR,ATTRIBUTES.GOODFORMEALBRUNCH VARCHAR,ATTRIBUTES.COATCHECK VARCHAR,ATTRIBUTES.SMOKING VARCHAR,ATTRIBUTES.DRIVETHRU VARCHAR,ATTRIBUTES.DOGSALLOWED VARCHAR,ATTRIBUTES.BUSINESSACCEPTSBITCOIN VARCHAR,ATTRIBUTES.OPEN24HOURS VARCHAR,ATTRIBUTES.BYOBCORKAGE VARCHAR,ATTRIBUTES.BYOB VARCHAR,ATTRIBUTES.CORKAGE VARCHAR,ATTRIBUTES.DIETARYRESTRICTIONSDAIRYFREE VARCHAR,ATTRIBUTES.DIETARYRESTRICTIONSGLUTENFREE VARCHAR,ATTRIBUTES.DIETARYRESTRICTIONSVEGAN VARCHAR,ATTRIBUTES.DIETARYRESTRICTIONSKOSHER VARCHAR,ATTRIBUTES.DIETARYRESTRICTIONSHALAL VARCHAR,ATTRIBUTES.DIETARYRESTRICTIONSSOYFREE VARCHAR,ATTRIBUTES.DIETARYRESTRICTIONSVEGETARIAN VARCHAR,ATTRIBUTES.AGESALLOWED VARCHAR,ATTRIBUTES.RESTAURANTSCOUNTERSERVICE VARCHAR,HOURS.MONDAY VARCHAR,HOURS.TUESDAY VARCHAR,HOURS.WEDNESDAY VARCHAR,HOURS.THURSDAY VARCHAR,HOURS.FRIDAY VARCHAR,HOURS.SATURDAY VARCHAR,HOURS.SUNDAY VARCHAR );

View για το table USER12.YELPCHECKIN:

CREATE VIEW USER12.YELPCHECKIN ( ROWKEY VARCHAR PRIMARY KEY,PERHOUR.BUSINESSID VARCHAR,PERHOUR.WEEKDAY VARCHAR,PERHOUR.HOUR VARCHAR,PERHOUR.CHECKIN VARCHAR );

DROP VIEW USER12.YELPCHECKIN;

SELECT \* FROM USER12.YELPCHECKIN LIMIT 10;

SELECT \* FROM USER12.YELPBUSINESS LIMIT 10;

Q1:

SELECT BASE.NAME,BASE.STATE,BASE.STARS FROM USER12.YELPBUSINESS WHERE BASE.ISOPEN = '1' LIMIT 1000;

Q2:

SELECT BASE.NAME,BASE.ADDRESS,BASE.CITY,BASE.REVIEWCOUNT FROM USER12.YELPBUSINESS WHERE BASE.CATEGORIES LIKE 'Drugstores' ORDER BY BASE.REVIEWCOUNT DESC;

Q3:

SELECT SUM(REVIEWCOUNT),CATEGORIES FROM (SELECT COUNT(BASE.REVIEWCOUNT) AS REVIEWCOUNT,BASE.CATEGORIES AS CATEGORIES FROM USER12.YELPBUSINESS WHERE BASE.ISOPEN = '1' AND HOURS.MONDAY = '0:0-0:0' AND HOURS.TUESDAY = '0:0-0:0' AND HOURS.WEDNESDAY = '0:0-0:0' AND HOURS.THURSDAY = '0:0-0:0' AND HOURS.FRIDAY = '0:0-0:0' AND HOURS.SATURDAY = '0:0-0:0' AND HOURS.SUNDAY = '0:0-0:0' GROUP BY BASE.CATEGORIES) GROUP BY CATEGORIES;

Q4:

SELECT COUNT(BASE.NAME),BASE.STATE FROM USER12.YELPBUSINESS WHERE ATTRIBUTES.SMOKING = 'TRUE' AND HOURS.SUNDAY != 'None' GROUP BY BASE.STATE;

Q5:

SELECT SUM(COALESCE(TO\_NUMBER(REGEXP\_SUBSTR(PERHOUR.CHECKIN, '^\d+(\.\d+)?')), 0)) AS CHECKINS, PERHOUR.WEEKDAY, PERHOUR.HOUR FROM USER12.YELPCHECKIN GROUP BY PERHOUR.WEEKDAY,PERHOUR.HOUR;

Q6:

SELECT SUM(COALESCE(TO\_NUMBER(REGEXP\_SUBSTR(PERHOUR.CHECKIN, '^\d+(\.\d+)?')), 0)) AS CHECKINS, BASE.CATEGORIES FROM USER12.YELPCHECKIN FULL OUTER JOIN USER12.YELPBUSINESS ON USER12.YELPBUSINESS.ROWKEY = USER12.YELPCHECKIN.ROWKEY WHERE BASE.ISOPEN = '1' AND PERHOUR.WEEKDAY != 'Sat' GROUP BY BASE.CATEGORIES;

AND PERHOUR.WEEKDAY != 'Sun'

WHERE BASE.ISOPEN = '1' AND PERHOURS.WEEKDAY != 'Sat' AND PERHOURS.WEEKDAY != 'Sun' AND PERHOUR.HOUR

Q7:

SELECT BASE.NAME,BASE.NEIGHBORHOOD,BASE.ADDRESS,BASE.CITY,BASE.STATE,BASE.POSTALCODE,BASE.LATITUDE,BASE.LONGTITUDE,BASE.STARS,BASE.REVIEWCOUNT,BASE.ISOPEN,BASE.CATEGORIES FROM USER12.YELPCHECKIN FULL OUTER JOIN USER12.YELPBUSINESS ON USER12.YELPBUSINESS.ROWKEY = USER12.YELPCHECKIN.ROWKEY ORDER BY PERHOUR.CHECKIN DESC LIMIT 100;